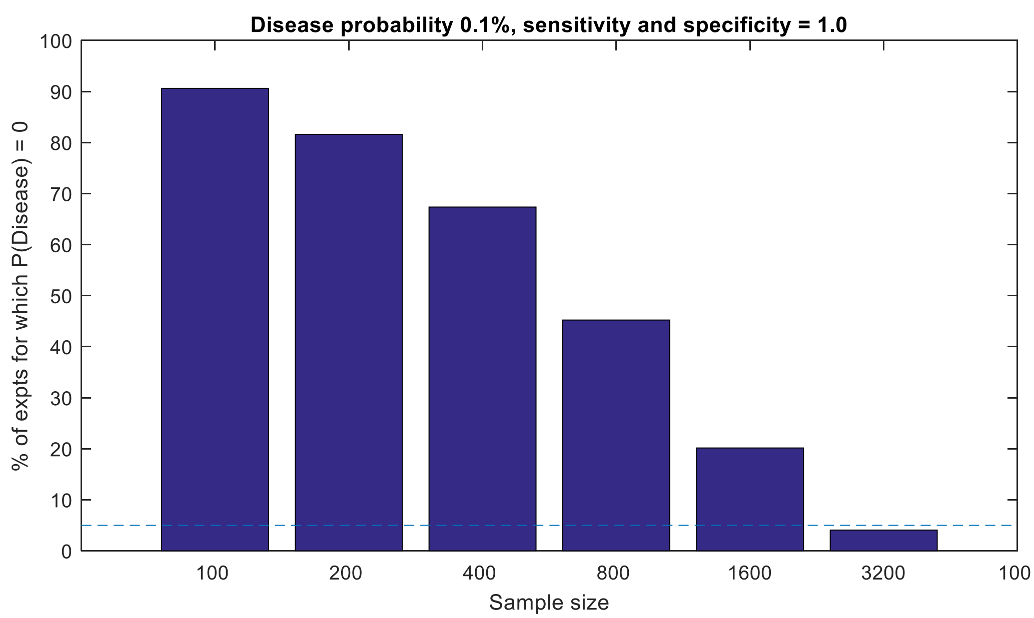
Supplementary Material

Disease incidence and statistical power

An important point requiring discussion is the statistical sensitivity of any particular study and its capacity to detect the presence of an effect within a population. This is generally referred to as the statistical power of a population sampling study and quantifies the capacity to measure an experimental effect and/or to determine the proportion of the underlying population that show such an effect.

If a population is assumed to be normally distributed with respect to a particular feature, then only a particular fraction of that population is likely to display such sensitivity above a certain threshold. In the case of possible health effects of wind turbines, the relatively infrequent reports of ill health effects suggest that the sensitive proportion of the population is quite low. The capacity to detect such sensitivity using a population sampling approach will be dependent on the underlying proportion in the population that display such a sensitivity and the number of samples taken from the population. For instance, for two relatively common clinical conditions of the inner ear, it is estimated that 0.2% of the population suffer Meniere's disease (Alexander & Harris, 2010) while recent estimates of superior semicircular canal dehiscence are around 0.1% of the population .

We have used a simple model of statistical power (see below) to examine the impact of sample size using the assumptions that (i) the incidence of the disease is similar to that of semicircular canal dehiscence (i.e. 0.1%) and (ii) for simplicity we have chosen a best-case scenario where the sensitivity of the test for the disease is perfect (100% detection and no false alarms).



*The figure plots the percentage of studies that would fail to detect the disease with an actual 0.1% incidence with sample sizes ranging from 100 to 3,200. The dashed blue line indicates 5%. The above model indicates that (i) for sample sizes of 400 or less there is 70% to 90% chance that the test would **fail** to return a positive result; (ii) a sample size of around 3200 samples is needed before there is a less than 0.05 chance that the test will provide a type II error (i.e. incorrectly retain a false null hypothesis). In this* *simulation there was a disease (0.1% rate) and the blue bars represent the % of tests that would return P(disease)=0 i.e. a type II error.*

This indicates that, of the many population studies of the effects of wind turbines on the population where the focus has been on random sampling, the numbers of samples (generally of the order 10^2) are well below the numbers that would be required to reliably detect the other relatively common inner ear conditions (order 10^4 samples).

In the absence of any other data, the disease incidence is estimated from the incidence of superior semicircular canal dehiscence (SCD: 0.1%). A systematic analysis of health practitioner records as a percentage of the total exposed population in potentially affected regions might provide a more grounded indication of the likely fraction of the population that may be susceptible to infrasound (presuming that is based on some similar inner ear dysfunction).

A second and critical issue in the analysis of such low incidence occurrences is how they are identified and currently treated in the analysis of the population data. It is statistically impossible for a small number of low incidence samples to have any meaningful impact on summary statistics (the mean, median or mode) of a population. The meta-analyses of previous studies of potential health effects of wind turbines have also used more traditional summary and linear regression models. More appropriate would be the identification of "outliers" for further examination for physiological or perceptual differences as outlined in the main body of this article.